frontiers in Virtual Reality

ORIGINAL RESEARCH
published: 21 September 2020
doi: 10.3389/frvir.2020.00011

Investigating the Perceived Strengths and Limitations of Free-Viewpoint Video

Andrew MacQuarrie and Anthony Steed*

Virtual Environments and Computer Graphics Group, Department of Computer Science, University College London, London, United Kingdom

Free-viewpoint video (FVV) is a type of immersive content in which a character performance is filmed using an array of cameras and processed into a video-textured, animated 3D mesh. Although FVV content has a unique set of properties that differentiates it from other immersive media types, relatively little work explores the user experience of such content. As a preliminary investigation, we adopted an open-ended, qualitative approach to investigate these issues. Semi-structured interviews were conducted with six immersive content experts, exploring the perceived strengths and limitations of FVV as a content type. These interviews were analyzed using inductive thematic analysis. We identified five themes during our analysis: they don't look real, but that's okay; they can really move; they don't connect with me; encounter, legacy, and truth; no technology is an island. Our analysis reveals a wide range of future research directions and provides insight into which areas may produce the most benefit in relation to the user experience. We discuss, for example, the potential impact of difficulties in supporting user engagement, aspects related to visual quality such as the importance of responding realistically to environment lighting, and tensions between visual and behavioral quality. The analysis also highlights the complex interplay of factors related to the content itself, such as performance style and the use of creative production techniques to reduce the impact of potential limitations.

Keywords: free-viewpoint video, virtual reality, augmented reality, content creation, VR, AR, XR

OPEN ACCESS

Edited by:
Doug A. Bowman,
Virginia Tech, United States

Reviewed by:
Bernhard E. Riecke,
Simon Fraser University, Canada
Pierre Bourdin,
Open University of Catalonia, Spain

****Correspondence:***
Andrew MacQuarrie
andrew.macquarrie.13@ucl.ac.uk

Specialty section:
This article was submitted to Virtual Reality and Human Behaviour, a section of the journal Frontiers in Virtual Reality

Received: *12 June 2020*
Accepted: *31 July 2020*
Published: *21 September 2020*

Citation:
MacQuarrie A and Steed A (2020) Investigating the Perceived Strengths and Limitations of Free-Viewpoint Video. Front. Virtual Real. 1:11. doi: 10.3389/frvir.2020.00011

1. INTRODUCTION

Free-viewpoint video (FVV) is a type of filmed content that supports viewing with six degrees of freedom (6DoF). Although FVV has been studied in academia for more than two decades (Kanade et al., 1997), cumulative improvements in bandwidth requirements and quality mean it is now practical for use in production contexts (Narayanan et al., 1998; De Aguiar et al., 2008; Vlasic et al., 2008; Starck and Hilton, 2003; Starck et al., 2005; Starck and Hilton, 2007; Collet et al., 2015). As a result, a number of commercial FVV studios have opened in recent years (e.g., 8i, nd; 4DViews, nd; Dimension Studio, nd; Volucap, nd; Jump Studio, nd; Metastage, nd), giving rise to a proliferation in immersive content produced using this approach. Despite FVV's long incubation as a technology, there has been relatively little work exploring the user experience of content produced using this technique. As FVV has a unique set of properties that differentiates it from other content types, it becomes critical to understand how these properties influence the experience. This allows us to form an understanding of what the advantages of the technology are for content producers, as well as identifying what impactful limitations remain to be addressed by researchers.

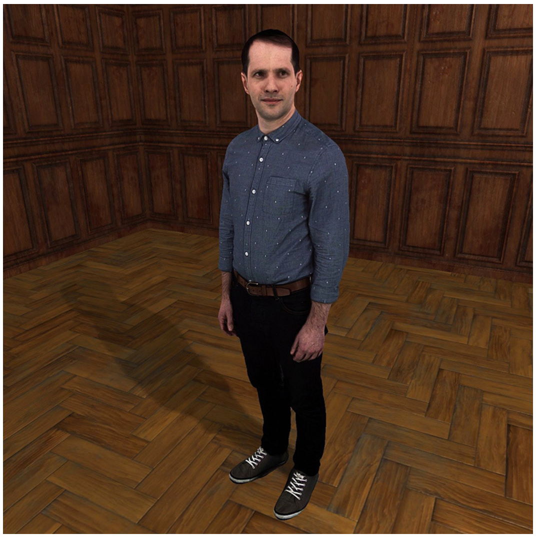

FIGURE 1 | An example of an FVV character mesh captured using a technique similar to that described by Collet et al. (2015), inserted into a virtual scene.

As so little is known about the user experience of FVV, we believe an exploratory investigation is the ideal first step. Provisional findings from this investigation can then be used to inform future research directions. To this end, we conduct one-on-one interviews with immersive content experts. This qualitative approach allows the generation of rich, open-ended data on individuals' experiences (Braun and Clarke, 2013). As FVV content may be viewed in a number of immersive display types, we often use the more generic term extended reality (XR) to capture the range of virtual reality (VR), augmented reality (AR), and mixed reality (MR) contexts in which it may be deployed.

In this paper, we focus on the current generation of commercially available, high-quality FVV content. There are a number of lower-end commercially available FVV techniques, such as DepthKit (Depthkit, nd), as well as more advanced techniques in use in academic settings that are not currently available for commercial use (e.g., Guo et al., 2019). As it is important to reflect on the current technology landscape, we reference these techniques where appropriate. It is important to note, however, that the analysis presented here cannot necessarily be generalized to all types and generations of FVV content. FVV content shown during our user study was created using a technique similar to that described by Collet et al. (2015). An example of an FVV character mesh captured using this technique and inserted into a virtual scene is shown in **Figure 1**. To allow readers to make a judgement about how the technology under discussion here compares to others we have made an example FVV scene, as well as a video of this scene, available on the project web page (MacQuarrie and Steed, 2020a).

In this research, we aimed to investigate the user experience of high-end, commercial FVV. The hope was to give insight into how this technology's unique set of properties influence the contexts in which it is perceived as being the most appropriate choice of content creation tool. Understanding these strengths may allow content producers to identify when they might find the technology most useful. Such an understanding may also allow researchers to identify ways in which these strengths can be more fully understood. Additionally, an understanding of the perceived weaknesses of FVV may allow researchers to focus on areas that are seen as being detrimental to the medium, and which would ideally be addressed in future generations of the technology. To this end, we formulate our research question as:

RQ: what are the perceived strengths and limitations of the current generation of high-end, commercial free-viewpoint video?

2. BACKGROUND

2.1. Filmed Content That Supports Viewing With Six Degrees of Freedom

FVV exists among an array of techniques for immersive media content creation. Indeed, a number of other methods have been proposed to allow filmed media to support viewing with 6DoF. In light-field capture, for example, camera views with a relatively small baseline are interpolated to create novel views (Levoy and Hanrahan, 1996; Zitnick et al., 2004). Other work has used texture synthesis to warp 2D views into novel perspectives (Xu et al., 2011). In contrast to these techniques, FVV creates an animated 3D mesh of a character performance, which is reconstructed from a live-action performance using an array of inward-facing cameras arranged around the actor (Kanade et al., 1997).

While commercial FVV studios differ to some degree in their approach, such as number and resolutions of cameras, many of the fundamental underlying concepts remain the same. A large number of red, green, and blue (RGB) cameras capture visible light, potentially used in combination with infrared cameras to improve surface reconstruction (e.g., Collet et al., 2015). Various techniques, such as green screen and depth-from-stereo are applied to each camera view to segment the actor from the background (8i, nd; Collet et al., 2015; Schreer et al., 2019). For each frame, shape-from-silhouette and depth-from-stereo techniques are applied to create a 3D mesh of the character. This 3D mesh is then textured using data from RGB cameras. This results in a video-textured 3D mesh of a character performance, in which appearance and surface dynamics are captured. Previous generations of FVV created a unique 3D mesh per frame, meaning FVV sequences were "temporally inconsistent" (Cagniart et al., 2010b, p. 333), which resulted in large bandwidth requirements (Collet et al., 2015). Techniques to reduce the volume of data are therefore applied, such as using a system analogous to P-frames in video, in which temporal similarities between frames are leveraged by encoding updates from the previous frame. In this technique, a temporally consistent mesh deforms over a relatively small number of frames before being replaced by a new mesh at the next keyframe (Collet et al., 2015; Schreer et al., 2019). This retains the robustness of unstructured meshes for portraying deforming shapes while reducing memory and bandwidth requirements to practical levels (Collet et al., 2015).

Recent academic works have demonstrated that it may by possible for volumetric videos to be edited and animated beyond the movement in the original recording (Regateiro et al., 2018; Eisert and Hilsmann, 2020). Currently, these techniques require fully temporally consistent meshes. It is not yet clear if temporally consistent meshes can provide the generalizability required for commercial FVV solutions. It has been proposed, for example, that "allowing dynamic mesh connectivity offers greater flexibility in capturing general scenes, e.g., with free-flowing clothing or interacting objects" (Prada et al., 2016, p. 108:2). To our knowledge, the current generation of commercial FVV content is designed to play out as recorded.

2.2. Other XR Content Creation Techniques

FVV content is filmed from a real character performance, capturing both surface dynamics and appearance. This sets it apart from motion capture (MoCap). In MoCap only the motion of a performance is captured, generally using infrared light and markers placed on the actor at specific locations (for a recent review of the literature, see Colyer et al., 2018), although alternative tracking systems such as inertial and magnetic are also available (Spanlang et al., 2014). This motion data can then be applied to a rigged avatar, where the animation of the mesh is driven through a skeleton structure (Baran and Popović, 2007; Spanlang et al., 2013). While this can create natural body motions for the avatar, the surface dynamics such as clothing are not captured by this technique, as motions are constrained by the prior articulated elements of the avatar, and surface appearance such as skin tone must be animated separately. Details such as mouth movement and facial expressions can also be added through lip synchronization or animation (e.g., Aneja et al., 2019; Gonzalez-Franco et al., 2020). One advantage of MoCap is that there are a variety of techniques for blending together separate MoCap sequences while maintaining plausible dynamics (Kovar et al., 2008). Techniques exist in academia to allow FVV clips to be blended together in real-time (Prada et al., 2016). To our knowledge such techniques have not yet been used in commercial content, and we have previously suggested that such techniques may currently be prohibitively expensive (MacQuarrie and Steed, 2020b). In practice, FVV content plays out as captured.

As FVV techniques result in a 3D mesh of a character performance, the technology gives viewers substantial freedom to move with 6DoF. Techniques such as light-field capture tend to be restrictive in the amount of motion they can allow, as views must be interpolated from a dense set of real-world camera locations (Levoy and Hanrahan, 1996). Likewise, a common technique to bring photo-realistic actor performances into XR is to use "billboarding" (Horry et al., 1997). In billboarding, cameras are used to film a character performance that is segmented from the background (e.g., using chroma-key techniques), with the resulting video displayed as a billboard inside the virtual space. This technique can produce 3D depth cues through stereoscopy, and some depth cues from head-motion parallax are available as the billboard moves relative to the 3D scene. Head-motion parallax cues within the character are not supported, however, and distortion increases as the user moves off-axis, until the character becomes completely flat when viewed at a 90-degree angle. Alternatively the billboard can rotate to face the user, although this can result in a character who appears to float around their central vertical axis, and this technique does not provide additional head-motion parallax cues. With FVV, as a user walks around the character they see that character from a different perspective, as in the real world.

Another common technique for VR content creation is 360-degree video. In 360-degree video, a video completely surrounds the viewer, and when viewed immersively it allows users to look around inside the scene naturally by turning their heads. While animated 360-degree content is available, a large amount is live-action. As live-action 360-degree video is filmed using a 360-degree camera, the content is photo-realistic, although low resolution can be an issue due to the high bandwidth requirements created by the extremely large field-of-view (Mangiante et al., 2017). 360-video, however, is fixed viewpoint i.e., only the three degrees of freedom related to orientation are supported. As 360-degree videos are generally not generated in real time, they usually play out as recorded. This property means 360-degree videos may be more suitable for non-interactive experiences (sometimes called "passive storytelling" Bucher, 2017, p. 67). FVV content may share this attribute to some degree. As FVV media currently plays out as captured, sections of content cannot respond to real-time events. FVV characters are 3D meshes though, so exist within real-time generated 3D environments, which may make it easier to integrate some "active storytelling" components. This may be similar to "choose your own adventure" style storytelling, which has been seen in 360-video as well (Dolan and Parets, 2016). In such techniques, input from the viewer can direct the narrative through a number of predefined paths.

2.3. The User Experience of Volumetric Video

A relatively small amount of work exists exploring the user experience of FVV content. Some work has been done on perception of FVV characters, specifically how accurately users can assess the eye-gaze direction of these characters. In their work on perception of FVV character's eye-gaze direction on screens, Roberts et al. (2013) explored how accurately users could perceive mutual gaze, i.e., when the FVV character was looking directly at the user. Their findings indicated the quality of the texture in the eye region impacted how accurately users could perceive its direction. In our previous work on gaze perception of FVV characters in head-mounted displays, we identified—in line with real-world findings—that users were much better able to assess eye-gaze direction when that gaze was directed nearer toward themselves (MacQuarrie and Steed, 2019). These works highlight what has already been identified as a potential issue in FVV: as the content is pre-recorded, the eye-gaze direction of the character cannot be altered in real-time to look at the user. There has been a great deal of work exploring the importance of eye gaze in human communication, and it is known to facilitate many aspects of social interactions such as conversational turn taking, communicating emotions and identifying objects of mutual attention (Argyle and Dean, 1965; Kendon, 1967; Argyle and

Cook, 1976; Novick et al., 1996). While it appears that eye gaze is therefore likely to be an important aspect of FVV, the impact on the technology as an XR content creation technique remains largely unexplored. Work has been done that explores the design space of techniques to ensure mutual gaze during volumetric teleconferencing (Anjos et al., 2019). In this work, point-cloud-based avatars were scaled dynamically to promote mutual gaze between conversational partners.

Work has also been done looking at the impact that a point-cloud-based volumetric representation of a human has on social presence during collaborative tasks, when compared against billboarded and avatar-based representations (Cho et al., 2020). Their results indicated that co-presence was highest for volumetric representations when the task required moving off-axis from the human representation, but that billboarding was as effective in eliciting co-presence when the task meant the representation was always viewed straight on. While this exploration of social presence is relevant to our work, their volumetric character representation was captured in real-time using a Kinect. This means their content type is not fully comparable to the offline, pre-processed FVV content under consideration here, as the set of properties inherent in the technologies are different (e.g., real-time versus pre-recorded, low-fidelity versus high-fidelity visuals, differing restrictions that the capture devices place on the performer, etc.). Also related to the user experience of point-cloud-based volumetric representations, recent works have explored the subjective experience of point cloud quality under different compression scenarios in AR (Alexiou et al., 2017) and VR (Subramanyam et al., 2020).

3. METHOD

As so little is known about the user experience of FVV, we believed it was important to start such an investigation through a relatively open-ended research approach. To that end, we adopted a qualitative research method. We used semi-structured interviews, conducted one-on-one with experts in the field.

Our research question intended to explore the perceived strengths and limitations of FVV. To that end we adopted an inductive approach within a contextualist/critical realist framework. We chose this method as we believe that individuals may perceive the technology and resulting media experiences differently, and potentially have differing opinions on their efficacy, but also that there are underlying features of the technology that will reflect in the data (Braun and Clarke, 2013). While the research techniques adopted here are not necessarily ideal for establishing generalizable truths, they are excellent for hypothesis generation on which further research can be based. Quantitative research methods, conversely, can have excellent properties around validation and reproducibility. However, they require structured hypotheses to test. Attempting to construct these hypotheses without foundation may introduce bias, eroding the validity of any findings made (Braun and Clarke, 2013). Through this work, we intend to aid future research by providing a grounding on which to base hypotheses.

Both authors are VR researchers looking to identify and explore the opportunities and most pressing concerns around FVV. We have both worked with FVV media ourselves, including pre-production, filming, post-production, and deployment. As such, we bring our own personal experience to this research.

3.1. Participants

Participants were approached whom the researchers believed would have insight into both XR content creation and the audience experience. The focus was on individuals who produce, commission, or evaluate XR experiences. The aim was to explore the content creation process and likely user experience, with people who had enough knowledge of the field to evaluate FVV against other content creation techniques, and who could articulate their opinions with reference to these technologies. As a result of these considerations, six interviews were conducted with experts in the field of XR content creation and evaluation. In this context, we use the word "expert" to mean someone who has been working professionally in the field for a number of years.

To ensure an open and honest discussion, participant anonymity was guaranteed. As the community is small, we provide only partial demographic data to avoid the risk of compromising this anonymity. All participants had a strong understanding of FVV, while four had first-hand experience of creating content of this type. Three participants were female and three were male. All participants were based in Europe. A stratified sampling approach was adopted, with participants recruited from different stages in the content creation pipeline. These stages were commissioning, production, and user experience evaluation. Participants were either identified and recruited directly through professional networks by the authors, or through snowball sampling.

3.2. Procedure

The study procedure was as follows. First, participants read an information sheet that detailed the experimental procedure, how their data would be used, and their right to withdraw from the study at any time. Critical aspects from this form, in particular the use of audio recording and the anonymization process, were discussed orally. Participants then signed a consent form. Audio recording was then started. Participants were then orally talked through the procedure, and informed of the aims of the research.

Participants were then invited to watch four FVV clips on a head-mounted display. Details of these clips will be discussed in section 3.3. While all participants had seen or worked with FVV media before, we felt it was important to show some content during the interview process. This allowed discussions to take place with reference to content known by both the participant and the interviewer. During interviews participants were encouraged to discuss FVV generally and to bring in knowledge and experiences from content they had previously worked with or seen.

Participants then took part in a semi-structured interview. These interviews were around 45 min long. The interviews were conducted by the first author, who roughly followed the interview schedule included in the **Supplementary Material**. This interview schedule was designed by the first author, and

TABLE 1 | FVV clip details.

Clip description	Character addresses user	Character approx. distance to user's initial position	Duration
A boy kicks a football around	No	1.5–2.5 m	15 s
A woman in a shirt and tie tells the user about sexism in the workplace	Yes	1.2 m	35 s
Two men break dance	No	1–3 m	27 s
A woman in period clothing makes an emotional appeal to the user	Yes	2 m	1 m 23 s

reviewed by the second author. The schedule was designed based on our own experience of FVV production and user experience, as well as issues highlighted in the literature and discussions that have been had within the XR community. The first participant acted as a pilot study to validate the interview schedule. No significant changes were made following this pilot, so we consider this participant to be part of the study. No significant issues were experienced during interviews, and no changes were made to this procedure as interviews progressed.

The study was approved by the UCL Research Ethics Committee (project ID 4547/012). Participants were given £10 compensation for taking part.

3.3. Materials

To allow the interview of busy professionals, it was required that the interview be conducted at times and places convenient to the participants. This meant ensuring the head-mounted display used to show FVV clips to participants was portable. As a result, the FVV clips were demonstrated on an Oculus Quest. While the clips ran smoothly on this relatively low-powered device, some compromises were made in other respects to the visual quality. The virtual scene in which the FVV clips were shown was very simple, showing an empty room that was generic enough that it was a plausible setting for each FVV clip. A fast type of anti-aliasing was used, specifically Fast Approximate Anti-Aliasing (FXAA). Most scene lighting was diffuse, which was explained visually in the scene by three large translucent windows. To allow the FVV clips to integrate more plausibly into the scene, a single overhead light allowed the FVV characters to cast a shadow on the floor. For graphical efficiency, the lighting was baked and the shadows low-resolution. As the meshes of the FVV clips are unstructured, changes in the topology can cause a visible flickering effect across surfaces when responding to environment light; as a result, the material was set to be purely emissive.

Four FVV clips were demonstrated. These clips accompany the software development kit distributed by Dimension Studio (nd). They were not created by us, but instead represent high-quality content being produced by a professional FVV studio. A summary of the clips is shown in **Table 1**. An example of an FVV character inserted into the demo scene shown to participants is shown in **Figure 1**. Note that, for licensing reasons, the example character in **Figure 1** and the demo clip available on the project web page are not identical to those shown to participants. They were created by the same FVV studio, however, so are representative. Participants were not given any specific instructions on how to engage with the content, and were told they could re-watch the clips at any time. Participants could walk within the confines of the play space (usually around 1.5 × 1.5 m), but could not locomote using the controller. FVV clips played out as recorded.

In the first of the clips in which the user is addressed by a character, the character was placed at a distance of around 1.2 m from the user. A distance of 1.2 m was selected because this is approximately the border between personal and social space, as determined by Hall (1966). The second clip in which the user was addressed was set at around 2 m. As the actor swayed slightly in this clip, the eye line is not uniformly directed. To compensate for this, we place the character at a slightly larger distance. For the two clips in which the user is not addressed, the distance between the user and the characters vary as these clips contain a large amount of character movement; the ranges for these clips are shown in **Table 1**. In all of the clips, the initial camera position was set vertically to be the FVV character's eye height. This was done to ensure vertical eye lines were correct. This means the virtual floor level was not necessarily consistent with the physical floor level, with this level of inaccuracy depending on the discrepancy in height between the FVV actor and the participant.

The position of the user at the start of each clip is set to a predefined position, such that the eye lines appear correct. However, the user is allowed to move with six degrees of freedom, so these eye lines are not guaranteed to work throughout. To encourage the eye lines to be as accurate as possible, the participant's position is reset between each clip. To allow this, the world fades to black at the end of a clip. The participant was told that when this happened, they should return to a neutral standing position and pull the trigger on the controller, which brought the scene camera to the ideal eye location for the next clip before fading from black back to the scene.

3.4. Method of Analysis

We follow a thematic analysis approach as outlined by Braun and Clarke (2013). Interviews were transcribed orthographically and data analysis was conducted on these transcripts. We do, however, "clean up" quotes for readability in this report. We remove hesitations, repetitions, and non-verbal utterances without acknowledgment, and introduce capitalization and punctuation as required for sentence structure. We use "[…]" to indicate any other removed text, while words in square brackets have been added.

Interview recordings were transcribed by the first author to facilitate data familiarization. Coding was performed inductively by the first author using nVivo 12. For practical reasons, some coding was performed as interviews progressed. A complete coding was performed across the first five transcripts by the first author. These transcripts were then reviewed by the first author, with codes systematically applied across the dataset. This resulted in 72 codes (codebook available as **Supplementary Table 1**). A sixth interview was conducted and coded. Another full review

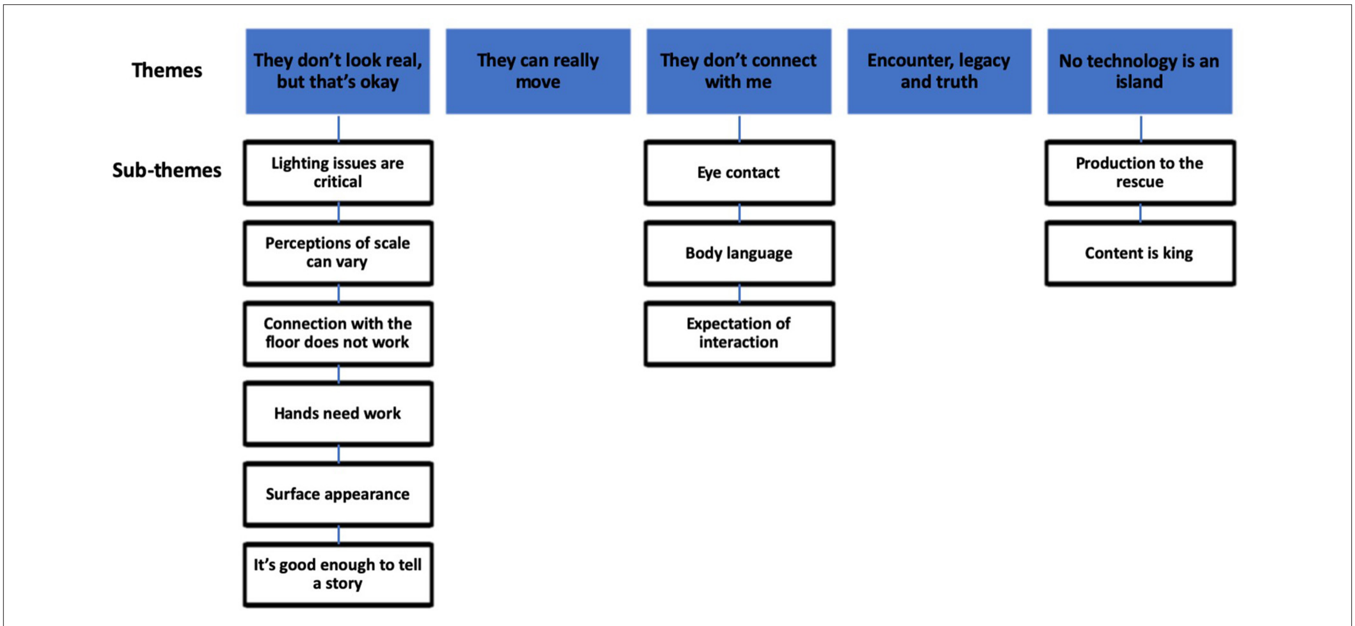
Themes
They don't look real, but that's okay
They can really move
They don't connect with me
Encounter, legacy and truth
No technology is an island
Sub-themes
Lighting issues are critical
Perceptions of scale can vary
Connection with the floor does not work
Hands need work
Surface appearance
It's good enough to tell a story
Eye contact
Body language
Expectation of interaction
Production to the rescue
Content is king

FIGURE 2 | Themes and sub-themes.

iteration of the transcripts and coding was then conducted by the first author. This resulted in 96 codes. Some grouping of the data into low-level related areas was performed to make the data easier to manage, resulting in a codebook with 99 nodes with some hierarchical structure (codebook available as **Supplementary Table 2**). These concepts were reviewed and grouped to generate candidate themes that were felt to capture the most important codes and concepts with respect to the research question. These candidate themes were discussed and reviewed with the second author. This process resulted in six themes (codebook available as **Supplementary Table 3**). During the write up, it was felt that two themes were in fact sub-themes of a larger theme, so were collapsed together. To ensure that the analysis accurately reflected the feelings of the participants, we performed member checking. A draft of the analysis was sent to participants, who were invited to provide feedback. All six participants said they were happy with their quotes and the analysis. One participant suggested additions in the analysis of Theme 4 (section 4.4, below). We highlight this change in the text when we come to it.

4. RESULTS AND DISCUSSION

Due to the broad range of issues raised by participants, we present discussion alongside results to improve readability. We identified five themes during our analysis. These were: they don't look real, but that's okay; they can really move; they don't connect with me; encounter, legacy, and truth; no technology is an island. A number of sub-themes were also identified. Themes and sub-themes are shown graphically in **Figure 2**. We attribute quotes using participant ID numbers throughout. As we believe value is not determined by comment frequency (Pyett, 2003), particularly with small sample sizes and fluid data collection techniques (Braun and Clarke, 2013), we favor less precise language when reporting comment occurrences.

4.1. Theme 1: They Don't Look Real, But That's Okay

The visual quality of FVV characters was generally felt to be good. All participants remarked that they were "impressed by the quality" (P4) in some respects, with the visuals described as "high-fidelity" (P1, P5). Despite this agreement that FVV produced high-fidelity visuals, there was also a feeling that "between real and not real, they're definitely in the not real space" (P2), and that an FVV character still does not "really look like a human" (P6). Here we see a clear distinction between a computer-generated image being visually impressive, and the successful illusion of reality XR may be striving to achieve.

There did not seem to be one specific aspect of the visual representation that participants highlighted as the cause of this lack of realism, but a number of issues were raised. Here we discuss the visual aspects that appeared to be the most important to participants in terms of the technology's ability to produce a realistic visual representation.

4.1.1. Lighting Issues Are Critical

Lighting was frequently mentioned as detrimental to realism: "It was strange because the level of realism was so high, and yet it didn't fit the lighting really" (P1). This reflects an issue in the current generation of FVV that employs temporally inconsistent meshes, in which the mesh changes every few frames (Collet et al., 2015; Schreer et al., 2019). When such FVV characters respond to environment light, a flickering is sometimes visible across the surface as the vector normals change rapidly. When P6 was discussing a recent piece that they themselves had made that contained this flicker, they felt that "half of the people are really happy about what we did, [while the other] half are not, because of that [flicker]," with some viewers feeling it looked "too strange" (P6). To avoid this flickering effect, FVV characters often do not respond to environment light. The examples that accompany the software development kit distributed by Dimension Studio, for example, rely on emissive textures (Dimension Studio, nd). This may have been responsible for one participant commenting that a character appeared to be "illuminated from within, rather than real-time," resulting in the character looking "like a light bulb" (P1).

As 3D reconstruction works best if the actor is diffusely lit on set (Collet et al., 2015; Schreer et al., 2019), volumetric pieces often fall back to diffuse environment lighting to ensure the character appears to fit plausibly in the scene (e.g., Alcon, 2017; Start, 2018). This may have an impact on the perceived tone of a piece:

> "Lighting did seem to me key [...] if you imagine that as your drama scenario, you would have dramatic lighting around it, and that felt really flat" (P4).

Despite difficulties in 3D reconstruction, some FVV pieces have chosen to light on set, resulting in dramatic lighting that fits the scene. An example of this is "Hold the World" featuring David Attenborough (Dodds, 2018; Sky UK Ltd, 2018). This was felt to have had a strong effect on realism, with one participant who had seen this piece commenting that they had felt "physically he was definitely there [...] sitting in that chair [...] he was very definitely physically there" (P2). It is for this reason that it was felt that "lighting on set is better," although this means that "you lose all the benefits of doing real-time animation" (P6), as the character cannot respond to real-time changes in environment light, as well as potentially causing issues with 3D reconstruction.

FVV clips being diffusely lit and not responding to environment lighting was felt to make the characters appear "very monoplane" (P1). One participant commented that "there is something about the solidity of the characters as well, the 3D nature of them, which doesn't feel completely solid," and while they were "not flat" they were also "not fully 3D [...] two-and-a-half-D or something" (P2). This loss of depth cues from shadows may have reduced realism, with one participant commenting: "somehow they still don't feel real in terms of depth and physicality" (P4). The impact this may have had on the user experience was not clear. While one participant commented that "the fact there weren't any shadows was eerie" (P1), another participant felt this lack of shadows only served "to undermine the sense of realism, but not to unnerve" (P2).

Allowing FVV characters to respond to environment light is an area that has received a substantial amount of research. One technique is to employ temporally consistent meshes, either by deforming a single mesh over the entire clip's duration (Ahmed et al., 2008; Cagniart et al., 2010a; Tung and Matsuyama, 2010; Huang et al., 2011; Budd et al., 2013; Mustafa et al., 2016) or by applying a known template mesh (Carranza et al., 2003; Loper et al., 2015). As the mesh is temporally consistent, the flickering effect seen as vector normals change rapidly is not present. Both of these techniques have issues with generalizability, however, as the selected mesh may struggle to represent varied or large shape changes (Casas et al., 2012; Collet et al., 2015). An alternative approach currently in development by Google combines FVV with lightstage techniques (Debevec et al., 2000), using time-multiplexed color gradient illumination to capture temporally consistent reflectance maps (Guo et al., 2019). While the results demonstrated using this technique are impressive, it represents an additional step up in complexity, and such techniques are not yet commercially available. These works do show, however, that future generations of commercial FVV may be able to address this issue. Based on feedback from participants, lighting appears to be a key limitation of the technology, as it may have a strong negative impact on the visual realism of the characters and how they integrate with the scene. This is likely to present particular issues for AR contexts, where lighting conditions cannot be known in advance but where an obvious goal might be to have the character fit plausibly into the visual scene around the user.

4.1.2. Perceptions of Scale Can Vary

The scale of the scene was frequently mentioned by participants. One participant felt the FVV characters and scene were at "the correct scale," commenting:

> "What I've felt in the past in many 360 videos is that they are out of scale [...] people in the scene, and especially if they come close they're like huge. [...] The proportions [in the FVV clips] felt right [...] in comparison to myself" (P3).

Other participants, however, commented that they had experienced issues with the scale of the scene, using phrases such as "I felt smaller than my physical self" (P1) and "I felt a bit like a giant" (P5). This likely reflects that the virtual camera height in the scene was fixed to ensure the eye lines of characters that addressed the user worked correctly. How inaccurate the virtual camera height was, therefore, depended on the height of the participant.

This is a complex issue, as our relationship to other characters in terms of height has multiple implications. One participant expressed expectations that were not met, saying they were "used to being taller than women," and indicated that this impacted their interpretation of the scene, commenting that it "felt weird, and set up a power dynamic" (P1). This issue of scale also impacted the sense of realism, with one participant commenting:

> "They look doll-like [...] that might be to do with the fact that I'm tall [...] they've been calibrated for another optimal height [...] so that feels unreal as well, so they feel like miniatures [...] I feel like I could almost pick her up" (P2).

As well as participants' perception of the character, mismatches in scale may also have had a negative influence on presence (i.e., the sense of "being there" Slater et al., 2009b, p. 193) and co-presence (i.e., the sense of "being with others" Zhao, 2003, p. 445):

> "Sometimes the height of the characters felt weird with where I was on the floor, so I didn't feel any sense of being in the space [...] at all with any of them" (P4).

The differing heights of actors has long been an issue in traditional filmmaking. Actors often stand on "apple boxes" to level heights and frame shots, with these boxes jokingly referred to as "man-makers" when used in this context (Uva, 2010, p. 37). Within immersive filmmaking this may present additional issues, as the whole actor is captured. The impact of camera height in 360-degree video is an open question (Keskinen et al., 2019; Rothe et al., 2019). In the context of MoCap performances, retargeting techniques are used to apply MoCap performances to avatars of a different size or shape to the actor (Hecker et al., 2008). While such techniques may not be directly applicable to the current generation of FVV, it is possible to edit the scale of characters in post-production, as FVV characters are standard 3D assets. In a piece one participant had recently worked on, they commented that they "play with [the FVV character's] size [...] deliberately" (P4).

Additionally, it was felt that when creating experiences it was important to consider users who may necessarily be at different heights, such as those less able to stand: "Imagine someone in a wheelchair for example [...] you'd be having them like float up in space" (P5). One participant commented that while it was "an artistic decision," designing "seat[ed] [experiences] can solve a lot of problems" (P4), such as helping reduce issues caused by differences in users' heights.

4.1.3. Connection With the Floor Does Not Work

In the FVV clips that we showed, a number of participants felt that "there's some issue about connection with the floor, it doesn't look realistic" (P2). One possible explanation given was:

> "The material and the finish of the environment and the shininess and the way the lighting reflects on the floor is different to how it reflects on the [character]" (P5).

This is an accurate assessment of the lighting issues mentioned previously, as the FVV characters do not respond to environment lighting correctly. This may indicate that this issue is particularly felt at locations at which the characters interact with the environment, where the contrast between expectation and rendered lighting may be particularly stark.

4.1.4. Hands Need Work

The visual representation of the character's hands was felt to be of critical importance, but also an issue in the technology that should be improved. For example, one participant felt that "the hands never work, the hands are really, really strange because you have little lines instead of fingers" (P6). One participant, however, commented that during the FVV clips we showed they had found the hands impressive:

> "I was amazed by the hands, because [...] having seen the stuff that's come out earlier from some of those volumetric things, the hands looked absolutely weird" (P4).

Getting the hands right seems critical. One participant commented that, given artifacts in the current generation of FVV, "if someone look[s] at the hands, it can be a little bit hard to get into the story" (P6), while another said of an early volumetric experience:

> "I really was quite distressed by some unworked volumetric stuff I saw [...] where the hands were like mutilated, horrific things, that really I haven't forgotten" (P4).

Commercial FVV tends to use saliency-based techniques to allocate additional mesh and texture quality to complex areas that users often look at, such as the face and hands (Collet et al., 2015; Schreer et al., 2019). As one participant commented that content in which hands were represented using fewer polygons was very badly received, this may indicate that this technique is helpful. It may not be sufficient, however, as other participants remarked that even in its current form artifacts in the hands caused issues. This may relate to difficulties in processing the capture rather than the allocation of resources in the mesh, as narrow objects such as fingers may be difficult to reconstruct (Collet et al., 2015).

4.1.5. Surface Appearance

Some comments highlighted FVV's ability to capture surface appearance, and in particular the accurate portrayal of skin tone. One participant remarked that "the skin tone is remarkably good" (P2), while another remembered noting during one example "how rosy [the character's] cheeks were, which I thought 'oh that's well done'" (P1).

One participant, however, felt that "the skin is the main problem" (P6) with the technology. Although they could not pinpoint exactly why they felt this, one given explanation was that the level of detail captured was not high enough. Another possible explanation is that skin is a famously difficult material to capture and render accurately, as real skin exhibits complex subsurface light scattering (Jensen et al., 2001). Although this issue is present in most real-time rendered XR content, high visual fidelity in other aspects could make the issue more pronounced in FVV.

4.1.6. It's Good Enough to Tell a Story

It may be that the perceived lack of visual realism is not necessarily a barrier to the technology's use. Although the current generation of FVV is "not perfect," it was felt that "it's enough to tell a story, it's enough to give emotion, it's enough to relate to a character" (P6). The content itself was generally felt to be the most important aspect.

When discussing visual realism, participants made frequent references to the "uncanny valley." The Uncanny Valley theory states that visual representations of people with near-human likeness can cause a stronger negative response than those with perfect human likeness or minimal human likeness (Mori et al., 1970). This theory "is a commonly cited, but controversial, explanation for human discomfort with imperfect human likenesses" (Mathur and Reichling, 2016, p. 29), in which the words "creepy" or "eerie" are often used to describe representations with near-human likeness . The theory is controversial because different studies into this effect have produced conflicting results (Mathur and Reichling, 2016). More recent investigations indicate that, although the causes are not fully understood, perceptual mismatches between artificial and human features may be responsible (for a review, see Kätsyri et al., 2015). Regardless, it is a commonly used term that is understood to indicate a creepy feeling produced by stimuli with near-human likeness.

Opinions varied on whether the current generation of FVV content produces an uncanny valley effect. When asked if the participants found the characters creepy in any way, four participants answered decisively: "no" (P1, P2)/"not really" (P3)/"personally I didn't" (P5). Reasons for this perceived lack of visual uncanny valley were mixed. One participant felt that "they were quite normal looking" (P2). Another participant felt that it related to their expectation:

> "It's because I acknowledged they weren't human, [...] they weren't real-time [...] so I think [that] my expectation was lower probably than if [it] had been a real-time thing" (P5).

One participant reported that FVV characters are currently perceived "not as human, not as cartoons, but as what they are, which is human avatars" (P6). If FVV characters currently have a similar level of realism as avatars, it may be that these characters are not yet "near-human enough" to be considered "uncanny." One participant, however, strongly felt that the characters produced an uncanny valley effect: "Is there Uncanny Valley here? [...] Yes of course [...] 100%," feeling that this was "because they pretend to be human, [and] it's strange" (P6).

The implications of issues in visual quality (e.g., lighting, hands, etc.) and their impact on the experience are not clear. While participants indicated that these aspects appeared to them to be detrimental to realism, there was also a feeling that realism may not be required for an XR experience to be successful. Previous work in the field of VR has explored the impact of visual quality, providing evidence that there is a complex relationship between realism and perceptual metrics. For example, while some findings indicate that increased self-avatar realism can lead to stronger body ownership (Gorisse et al., 2019) and increased scene realism can lead to increased presence (Slater et al., 2009a), other work did not find that increased character realism resulted in an increase in presence (Vinayagamoorthy et al., 2004). We believe there is substantial scope for future research here, exploring these visual aspects and their impact on the user experience.

4.2. Theme 2: They Can Really Move

All participants agreed that the technology's ability to portray "natural movements" was one of the clear "upsides of [...] volumetric capture" (P3). This reflected in participants' preference for clips that focused on movement:

> "The football and dance ones [are] the most compelling, because that [is] what it's brilliant at capturing, [...] that physical form" (P4).

It was felt that "the fluidity in the dancing is really interesting, [...] really effective" (P2). Indeed, participants felt that "the nature, the physicality of" the captured motion meant that the clips involving movement "were the most realistic" (P1). This feeling that FVV captures natural and realistic motion was highlighted by the types of performance participants felt were a good fit for the technology. Use cases that make use of the technology's accurate portrayal of motion were mentioned, e.g., "sports" (P3, P6) such as "wrestling" (P1), and arts such as "dancing" (P2, P4, P5).

As discussed in the Background, the portrayal of secondary motions such as clothing is one of the main benefits volumetric capture brings over MoCap applied to avatars. One participant found this component critical:

> "The quality of the clothing was really compelling and delightful in a way, you know, it was really beautiful. [...] The way it naturally flows, [...] you don't get that from [rigged] avatars." (P5)

While a general feeling of naturally flowing character motion was echoed by all participants, no other participants attributed it to the cloth dynamics. Indeed, some participants did not notice the clothing motion, feeling that it was "not such a big deal" (P3). Again this may have been due to the fact that FVV clips do not respond to environment lighting, meaning self-shadowing cues from folds in the clothing are missing:

> "Lit in a different way you might be more aware of the way the cloth and stuff moves, but I wasn't very aware of that" (P4).

This may indicate that one of the technology's key attributes in contrast to other techniques such as avatars is being undermined by limitations in FVV's ability to respond to environment light. Comments from participants also indicate the impact of secondary motions on the user experience are not clear, and this may warrant further investigation in the context of FVV.

4.3. Theme 3: They Don't Connect With Me

Participants reported feeling a difference between clips in which the character addressed the viewer and where they did not:

> "There's a distinct difference between some of the clips [...] like doing their thing [...] upping the ball or dancing, and then there's clearly the other two clips where the actor addresses you as if you were in the room" (P3).

For clips that addressed the user, participants "felt disconnected, because [the character] didn't react to [them]" (P5). Indeed, it was felt that "there was no acknowledgment that [the participant was] there" (P5). As with visual realism, there was not one aspect of the technology that participants attributed this loss of connection to. Here, we discuss a number of issues raised by participants that may have contributed.

4.3.1. Eye Contact

All participants agreed that a lack of accurate eye contact was a substantial issue when being addressed by characters. One participant even felt that "eye contact was the biggest flaw" of the technology (P1). This stems from the fact that FVV clips usually play out as recorded, so the eye gaze of the character reflects the gaze of the actor when it was filmed. This results in "a very tiny window where your eye contact matches," which causes issues when "trying to [...] take advantage of looking around, and the full volumetric nature of it," resulting "ironically" in "volumetric [content] that only works from a fixed position" (P1).

Participants felt that a lack of eye contact damaged the sense of connection with the character, as "you don't get that sense of somebody really engaging with you [...] even although they're right next to you," and that this may negatively impact their "emotional engagement" (P4). One participant felt that as "the eyes are not exactly looking at you, it [...] felt like [the FVV character was] looking at [...] a group of people behind me" (P5). If a user believes the FVV character is looking past them to another entity in the room, this may affect a user's understanding of the scene, and even cause concern that a person is standing behind them.

As mentioned previously, the height of an FVV character is a complex issue. FVV character height presents particular issues in situations where they look at the viewer. In order for eye lines to work, the actor must direct their gaze toward the viewer's head position. If the actor is taller than the viewer, the actor must look downwards, while if the actor is shorter, they must look upwards. This is dependent on the viewer height, which varies between users, and will only work when the user is at a specific distance from the character. As a result, FVV actors who address the user often direct their gaze at their own eye level, so the eye line is parallel to the floor and their distance from the user is no longer critical to ensure mutual gaze. This does, however, require the actor and viewer to be at a similar eye level, which may cause issues with perceptions of scale (see section 4.1.2 for further details).

As mentioned in section 2.1, current FVV content tends to play out as recorded, but research suggests in the future it may be possible to support real-time animation beyond the original recording. Finding a way to provide real-time eye gaze correction for volumetric videos may prove critical for their success in contexts where the user is addressed, as comments from participants indicate that this may be a serious limitation of the technology.

An additional issue with the eyes beyond mutual gaze could be problems with vergence, as the actor's eyes may not converge at the viewer's head position. One participant reported "trying to figure out if [the character was] cross-eyed or not" (P2). This issue may also result in confusion over eye-gaze direction, if the eyes do not appear aligned (MacQuarrie and Steed, 2019).

4.3.2. Body Language

Although eye gaze was seen as critical, participants felt this lack of acknowledgment from the characters went deeper. One participant noted that in real-world conversations, "when you look at someone it's not [just the] gaze [...] it's all [of the] body" (P2). This lack of acknowledgment in body language lead one participant to feel "like a spectator viewing the content, more so than an active part of the interaction" (P5).

Participants also commented on a lack of social mimicry. Social mimicry is the phenomenon in which humans tend to unconsciously mirror certain aspects of other peoples' behavior during interactions, such as posture and mannerisms (Chartrand and Bargh, 1999). One participant noted that when they "moved a bit closer [to the character] [...] they [did not] echo" (P1), while another participant felt the lack of small gestures of acknowledgment, such as "leaning in to the conversation" (P2). This led one participant to comment: "it feels like you're the only active agent in an otherwise dead scene of things that are on rails" (P1).

It has been shown in previous work that improving certain aspects of behavioral realism in avatars can improve to what extent users accept them as real. Realistic turn-taking eye-gaze patterns, for example, have been shown to increase measures such as co-presence, when compared against random gaze patterns applied to avatar-based conversational partners (Garau et al., 2001). Other nonverbal behavioral patterns may also prove to be important. The existence of social mimicry is well documented, and it has been shown to be an important component in interpersonal communications that can improve aspects such as positive judgements of others (Gueguen et al., 2009).

It is likely to be impossible to pre-record content that always follows these principles, as aspects such as the posture of individual users will impact what behavior a character must exhibit in order to be perceived as "behaving realistically." This may present a particular issue for FVV characters, as they currently tend to play out as recorded and therefore cannot easily respond to real-time events. Such mimicry behaviors are nuanced, however, and are not currently fully understood (Salazar Kämpf et al., 2018). As a result, a lack of such subtle gestures of acknowledgment is not only an issue in FVV, but still presents challenges in other forms of XR content (e.g., avatars), despite their ability to respond to events in real-time (Hale et al., 2020). As a result, this perceived limitation of FVV may not be as different from other forms of XR content as might initially be anticipated.

4.3.3. Expectation of Interaction

Participants wanted to be acknowledged by the character, but this desire was influenced by expectation. One participant commented:

> "[The characters] were intending to deliver an intense message, but it didn't feel intense on my side. [...] Probably it is because I immediately knew they were pre-recorded and they did not react to any of my [actions]. I knew it wasn't real time, [and that] they weren't there with me." (P5)

Participants had different feelings on what the impact of expectation might be for users. One participant felt that because FVV is "a new kind of [...] medium, [users] accept [...] the rules of engagement" (P6). Another participant, however, felt that the high-fidelity nature of the visual representation might lead to higher expectations of behavioral realism:

> "I think that's the problem here—that the fidelity is so high but the interaction is so low. [...] and in a way [... it's the uncanny valley, because it looks so good [...] [but] it breaks so easily." (P1)

Potential issues created by a discrepancy between visual fidelity and interaction have been noted previously in VR research. Indeed, research has suggested that "there is a need for consistency between the fidelity of the avatar's behavior and its appearance" (Garau, 2003, p. 6). It may be that higher visual quality creates an expectation of higher behavioral quality, and that this contrast may negatively affect how users perceive the avatar when that expectation is not met (Garau, 2003; Bailenson et al., 2005). The interaction between visual and behavioral realism remains unclear, however, with a recent work showing that more realistic facial animations led users to accept a self-avatar as themselves more readily, despite its cartoon-like appearance (Gonzalez-Franco et al., 2020). If a mismatch between high visual fidelity and low behavioral realism does negatively affect avatar perception, this may present a particular issue for the current generation of FVV. This is because FVV characters are perceived as having high visual fidelity, but potentially low behavioral realism due to their inability to respond to real-time events.

4.4. Theme 4: Encounter, Legacy, and Truth

Media experiences that capture a real person were mentioned as a strong use case for FVV, as "you feel like you're having a direct encounter" (P4). This is reflected in some of the types of experiences that are being commercially produced, such as "Hold the World" featuring David Attenborough (Sky UK Ltd, 2018) and "Wimbledon: Champion's Rally with Andy Murray" (Dimension Studio, 2019).

The reason participants felt FVV was appropriate for such experiences was due to the accuracy with which the technology portrays the captured person. Visual accuracy was mentioned, as "people need to recognize them [...] because [...] you've already seen their faces" (P6). The accuracy with which FVV portrayed physical movements was also highlighted. One participant gave the example of an FVV character that had captured "his stance correctly - he's got a very specific gait and posture," but also that more generally FVV can "capture the funny little [...] gestures" that characterize a person and "their true, natural, eccentric, idiosyncratic selves" (P2). Together, accurate capture of the visual and dynamic aspects of a person were felt to bring "more of a sense of that real person and what they're like," for example you could "it across the table from the real David Attenborough [...] feeling that it is him just as we [...] know him from film and television" (P4). This accurate portrayal of a real person was also given as a reason why FVV would be suitable for "the legacy thing, about being able to capture your family [...] and preserve it for some future" (P2).

Opinion differed on whether FVV was the only immersive content type suitable to portray such encounters. Participants largely agreed that it "makes an awful lot of sense to be using volumetric rather than avatars" (P2), as "an avatar of someone is more of a caricature of them [to] some extent" (P4). In relation to an encounter piece one participant was working on, they commented:

> "You get a direct encounter with [...] that person [...] so an avatar would not have worked at all for that, and then the question was, could a stereo pair of RED cameras or something [have] achieved the same result" (P4).

In this case, the "pair of stereo RED cameras" is a reference to billboarding techniques, with RED manufacturing high-end cameras for film and television. Opinions on the impact of the loss of support for head-motion parallax that billboarding suffers from were divided. One participant commented that for users comparing billboarding "verses volumetric [...] I think you'd be astonished at how few people would notice the difference" (P1). Another participant, however, felt:

> "Just having only the parallax [...] is changing everything [...] even just breathing [...] and seeing that the view is changing [...] even if you sit, 6DoF is amazing and much better" (P6).

Participants felt that a potential strength of FVV lies in the fact that it is a reproduction of a recorded event, and that this related to "a question about truth" (P2). Participants noted a potential difference in requirements between content types, with documentary and news following a different standard from fiction:

> "From an [...] ethics, production standard [perspective], it is that you've captured that real person [...] certainly in news and documentary [...] there is a sense of integrity to this, this is that person. [...] Once it's an avatar you get into reconstruction, [...] you get into that sort of fake news type thing. Is it really them? Whose version of it is it?" (P4).

One participant noted that an emotional documentary piece that they had made previously using MoCap had prompted a number of users to feel "betrayed" (P6) when they discovered the MoCap elements had been performed by an actress rather than the subject who was providing the voice-over. This experience prompted the participant to comment:

> "In documentary, truth can be important for the audience. I think in fiction [it doesn't] matter [...] we're all lying, we're all magicians" (P6).

In the field of immersive journalism, questions around relay "truth" through VR have been under debate for over a decade. In a seminal work in this area, it was posited:

> "Immersive journalism does not aim solely to present the facts, but rather the opportunity to experience the facts. We stress that the distinction between conventional documentary content, such as video and audio recordings, and synthetic content, such as 3D models and animation, is blurring." (De la Peña et al., 2010, p. 299)

This prediction of a blurring between "conventional documentary" and "synthetic content" seems particularly relevant in this context, with FVV potentially possessing qualities of both video and 3D models. This may indicate another layer in FVVs ability to deliver some "truth," as it presents not only an event as it was recorded, but may also provide viewers with the opportunity to experience that event as if they were there. During member checking, however, P4 noted that they believe journalism is unlikely to be a common use case for the current generation of FVV. They felt that the high cost of production was prohibitive for most journalistic budgets and that, due to the nature of the studio filming required, all FVV content was inherently a performance. Research from the field of VR journalism highlights the potential importance of "truth," however, that may be important for encounter and legacy purposes, which were raised as possible use cases for the technology.

If FVV is perceived as providing some guarantee of "truthfulness," this potentially carries additional ethical considerations. This may become more acute as the realism of FVV characters improves. In a recent work on the ethical implications of increased visual realism in XR, Slater et al. (2020) underscored the importance of following good practice from other media types, such as clear content warnings and the use of editorial guidelines, although they highlight that the details of how these may apply in XR are yet to be determined. Feedback from participants in our interviews indicates this may be an area that warrants further investigation.

4.5. Theme 5: No Technology Is an Island

A large portion of input from participants did not just focus on the technical aspects of FVV, but took in a broader view of the content and the processes involved in producing it. This wider ranging discussion indicates that the technology cannot be considered in isolation. Many aspects of a piece such as the writing, acting, environment, post-production, etc., will play into how the strengths and limitations of the technology are experienced by users. This may imply that future research could benefit from adopting a more holistic approach, considering not just the technical aspects, but also the implication of how these will likely interplay with other factors. Here, we discuss some of these factors, and how participants felt they related to the strengths and limitations of the technology.

4.5.1. Production to the Rescue

Participants agreed that every medium has its limitations. Creatives work around these limitations through artistic treatments "like magic tricks" (P6); distracting from, disguising or utilizing these limitations to produce pieces that work. Such techniques are "what cinema is doing [...] all the time" (P6), and participants felt that it may be possible to adapt some "tricks that we know from other media" (P2) to improve FVV content.

An example of such a technique was to design content in which the character was at a greater distance from the viewer. One participant proposed a distance of "2 m minimum," as users will not "see all the details [they] would have seen at 1 m" (P6), meaning fidelity issues and artifacts are disguised. Lighting was also mentioned, with participants noting that "dark is better" (P1) to disguise artifacts, as the user "cannot see what is wrong" (P6). The use of costume was discussed, with one participant suggesting the use of sunglasses for "obscuring [the loss of] actual eye contact" (P1), although another felt that "sunglasses [don't] work, it just removes you too much" (P4).

Real-world commercial examples of FVV content provide some indications of how creatives have worked around limitations in the technology. One participant noted issues when trying to "transition seamlessly from one [FVV clip] to another" (P3) within a scene, making interactive narratives difficult to construct. In "Blade Runner 2049: Memory Lab" (Alcon, 2017), an FVV character is framed as a hologram, so their sudden appearance and disappearance to allow cuts between clips is integrated into the narrative. The flickering effect when FVV characters respond to real-time lighting is disguised using the flickering of an open fire in "Awake: Episode One" (Start, 2018). In a recent piece a participant had worked on, they reported using distraction techniques by animating a very dynamic scene, ensuring the character is "not your only focus" (P4). Likewise, additional treatment on the piece such as "putting a volumetric character into a slightly imaginary situation" may help, as "you suspend your disbelief for it, because you sort of know [the character is] not real" (P4).

In what was considered to be a major barrier of the technology, all participants agreed that creating FVV content is expensive. In particular, processing the multi-view videos into a 3D mesh is costly due to the processing time required (MacQuarrie and Steed, 2020b). As a result, reducing the duration of FVV content reduces costs. In one piece a participant had worked on, the FVV character only appears for a few seconds in each section, before they fade out and the rest of the dialogue is conveyed via voice over.

From these comments it is clear that creative treatments are often used to disguise weaknesses. It may be important in future work, therefore, to consider strengths and limitations as being affected by context and content, and taking a more holistic view of the ways in which the technology can be used.

4.5.2. Content Is King

Various aspects of the content itself were felt to be critical for its success. The framing of the piece, various qualities of the acting, the scenery and setting, as well as how the user perceives themselves, were all felt to be important considerations when making an experience using FVV characters.

A frequent comment was that participants felt that they were "watching a performance, rather than being part of a performance" (P5). While a lack of acknowledgment was felt to contribute to this feeling, qualities of the acting were also highlighted: "it was the lack of eye contact, it was overblown gestures, the tone of the voice—it felt very monologic" (P1).

Different mediums require different performance styles (Bernard, 1997). It was felt that FVV production may currently be "not unlike early cinema, [where] the acting had to adapt to what could be captured" (P4). One participant felt that FVV might be similar to "directing for games," which requires "a slightly more physical performance," although they felt that FVV might benefit from a "voice [that's] very naturalistic [. . .] a very intimate voice" (P4). Participants agreed that the example clips we showed did not capture this style of acting. One participant commented that during the clips it was "not as if someone [was] talking to you personally [. . .] there's no [. . .] intimacy" (P2), while another felt the acting was "very hammy" (P1). It was felt that "technical constraints" placed on the captured person—such as the requirement to move fingers and arms in particular ways, such as to avoid self-occlusion—may be "straitjacketing [the performance] a bit" (P2). While participants felt that "actors [can] do that, that's their job," it may lead them to "go into acting mode," which could mean "that you lose naturalism" (P2). Additionally, it was felt important to consider non-actors:

> "A lot of volumetric capture is used to shoot non-actors [. . .] performers, dancers, footballers, [and] these people are not used to it [. . .] and you can tell them 20 times [. . .] not to do that, [and they] will still do that, [. . .] because [they are] not an actor" (P6).

This loss of naturalism in the performance may have contributed to participants feeling that they were watching a "stage play" (P1). This recurring perception that participants were watching a play may give rise to issues around how viewers perceive themselves during these pieces. Our demo scenes did not feature a virtual body for the viewer, which may have contributed to this issue. Users feeling like a "floating bodyless camera" (P1) may make them "not feel part of it" (P5). This may produce "an implicit asymmetry," because the FVV characters "had all their clothing, they were pretty high fidelity, and [the viewer] was nothing" (P5). While a lack of embodiment may have contributed to this issue, it was also noted that a self-avatar may even make things worse, if "it reinforces the fact that [the FVV characters] should be able to see you but yet they don't react to you" (P5).

A number of aspects of an experience may cause a user to feel like a "ghost," with Bye (2017) proposing that the inclusion of a self-avatar, the level of agency users have, and the extent to which their presence is acknowledged in the scene all contribute. Indeed, the impact of being acknowledged or not by characters has been discussed before in the field of VR storytelling. Not being acknowledged was termed "The Swazye Effect" by Oculus Story Studio, who described this sensation as "having no tangible relationship with your surroundings," which they felt "can create a considerable gap in connecting with the story and action" (Oculus Story Studio, 2015). This supposition has some quantitative support. A recent work by Steed et al. (2018) found a significant improvement in presence and experiencing the situation and characters involved as real when the user is subtly acknowledged by avatars in the scene through responsive eye gaze. While this study also found evidence that the presence or absence of a self-avatar impacts illusions of reality, it was to a smaller degree then being included in the scene through being acknowledged. In our study, FVV characters' failure to successfully acknowledge the user was felt to be significant, with one participant commenting:

> "The only way to make [the experience] okay is to present a scenario where it doesn't feel so obvious that [the characters are] not acknowledging you" (P4).

One participant felt it would be better to position the viewer "as part of an audience," because users would then "be more tolerant to [. . .] non-interaction" (P5). Others felt that the actor's performance could be used to set the expectations of the user, with one participant commenting of an encounter piece: "I'm not going to ask him questions [because] he's just telling me a story" (P4).

From these comments it is clear that many aspects of the context of a piece will affect users' expectations, and therefore the impact of technical constraints. Setting up these expectations through the staging, the performance style of the actors, or the presence or absence of a self-avatar, are all areas raised by participants that it would be interesting to explore further. Additionally, it may be important to consider these factors when assessing technical developments to the production pipeline, as there is likely to be substantial interplay between these facets.

5. LIMITATIONS

Our approach of collecting data from XR experts using flexible qualitative methods has a number of limitations. The use of experts means we are capturing a specific point of view that may not represent the experience of the average XR user. It also means participant numbers were low, meaning it is not possible to claim our findings are generalizable or complete. As a result, all findings here must be considered preliminary and require further investigation.

Our choice of the Oculus Quest as a device to demonstrate FVV content to participants also had limitations. As discussed in the section 4, a number of issues may have arisen from this. Issues such as the resolution, aliasing, and the need to simplify the scenery, may have affected participants' perceptions of the media. All participants were familiar with FVV or had worked with it extensively, however, and these clips were intended to act as a shared touch point to facilitate discussion. Additionally, the Quest is one of the most popular VR device currently on the market (SuperData, 2020). As a result, this reflects a common use-case for the technology. Physical space restrictions during clip demonstrations were also highlighted by participants during interviews, as an ability to move more freely may have allowed a different type of engagement with the FVV characters, such as to approach and examine them.

The coding was performed by the first author only, which could have impacted the analysis and resulted in an interpretation that is constrained to a single viewpoint. To mitigate the impact this may have had, the analysis was reviewed by both authors, and member checking was utilized to ensure participants felt their views had been faithfully interpreted and that our corresponding analysis was appropriate.

6. CONCLUSION

FVV has recently become practical as a content type for production XR material. FVV has a distinct set of properties that set it apart from other XR content types. It is therefore essential for us to form an understanding of how these properties impact content creation and the user experience, and this may help to guide future research.

We felt it was important to start this investigation in an open-ended way. With this aim, flexible qualitative data collection and analysis techniques were employed. Six experts in the field of XR content were interviewed. These interviews were then analyzed using inductive thematic analysis. The result was a wide-ranging analysis and discussion that covered a broad spectrum of potential strengths and weaknesses in the current generation of FVV. This analysis resulted in five themes:

1. They don't look real, but that's okay
2. They can really move
3. They don't connect with me
4. Encounter, legacy, and truth
5. No technology is an island

In the first theme participants identified that in terms of visual quality, although FVV characters were considered high fidelity, they still do not appear "real." Despite this, it was felt that a lack of realism may not necessarily have a negative impact on the user experience, and the technology was felt to be good enough already for many use cases. The content's current inability to respond correctly to environment light, however, was seen as a technical issue with wide ranging implications, such as impacting scene integration, perceived solidity and depth, and reducing the effectiveness of secondary motions such as cloth dynamics. The impact of scale was also found to be significant, and issues caused by differences in height when an FVV character addresses the viewer were felt to have implications on presence and scene understanding. Improvements in visual quality were also felt to be important, with surface appearance and hands mentioned as two particular areas where increased quality could be important.

In the second theme, accurate portrayal of natural movement was generally agreed to be a strength of the technology in comparison to other XR content creation techniques. In the third theme, however, a lack of behavioral realism beyond this movement was identified as a limitation. FVV characters that address the user were felt to present a particular issue, with unrealistic eye contact and body language mentioned as key factors. While increased visual realism is laudable, we discussed evidence in the literature that such improvements may increase expectations of behavioral realism. Investigations into the impact of these factors may prove essential as the gap between visual and behavioral realism widens. Additionally, while we discussed that recent research may help address issues in behavioral realism, in the fourth theme it was identified that this may potentially be in conflict with the perception of the technology's ability to reproduce an event faithfully. FVV's ability to accurately portray real people and events was seen as a key strength over other technologies such as avatars. This strength, however, may also give rise to important ethical considerations around the portrayal of "truth," with these considerations potentially being contingent on the content type.

In a fifth theme, it was highlighted that all mediums suffer from some limitations, and that creatives can work around these to produce awe-inspiring content. These production considerations, such as using tricks to hide or even utilize limitations, as well as the complex interplay of factors such as writing, acting and post-production, are critical components of the FVV pipeline. These elements will have a large impact on to what extent users notice limitations of the technology. Investigations into these factors from producers, researchers, and commercial studios would be highly useful, as it appears there is a large amount of scope for improvement within the current generation of FVV. Commercial studios in particular are likely to be generating invaluable experience about performance style, production techniques, etc. While such methods may be commercially sensitive in the short term, we would encourage the open sharing of findings to allow the technology to thrive. This could help establish "best practice" in the field, and facilitate a full comparison of FVV to other content creation techniques, which we believe could be ideal routes for future work.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because they are qualitative interview transcripts. Providing these would carry the risk of revealing the identity of our participants. As such, we do not make them available. The authors are happy to answer specific questions on the dataset, such that does not damage participant anonymity. Requests should be directed to: andrew.macquarrie.13@ucl.ac.uk.

ETHICS STATEMENT

This study was reviewed and approved by UCL Research Ethics Committee (project ID 4547/012). The participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

AM and AS conceived and designed the study and reviewed the analysis. AM conducted the interviews and performed data transcription and analysis. AM wrote the paper with the help of AS.

FUNDING

This work was supported by grant EP/N509577/1 from the UK Engineering and Physical Sciences Research Council (EPSRC).

ACKNOWLEDGMENTS

The authors thank the actors Joseph Chance and Sarah-Louise Young, whose likenesses appear in the figures and supplemental material of this work.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frvir.2020.00011/full#supplementary-material

REFERENCES

4DViews (n.d.). *4DViews - Volumetric Capture Systems*. Available online at: https://www.4dviews.com/ (accessed May 14, 2020).

8i (n.d.). *8i*. Available online at: https://www.8i.com/ (accessed May 14, 2020).

Ahmed, N., Theobalt, C., Rossl, C., Thrun, S., and Seidel, H.-P. (2008). "Dense correspondence finding for parametrization-free animation reconstruction from video," in 2008 *IEEE Conference on Computer Vision and Pattern Recognition* (Anchorage, AK), 1–8. doi: 10.1109/CVPR.2008.4587758

Alcon (2017). *Blade Runner 2049: Memory Lab*. Available online at: https://www.oculus.com/experiences/rift/1789924451050066 (accessed May 15, 2020).

Alexiou, E., Upenik, E., and Ebrahimi, T. (2017). "Towards subjective quality assessment of point cloud imaging in augmented reality," in *2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP)* (Luton, UK), 1–6. doi: 10.1109/MMSP.2017.8122237

Aneja, D., McDuff, D., and Shah, S. (2019). "A high-fidelity open embodied avatar with lip syncing and expression capabilities," in *2019 International Conference on Multimodal Interaction*, (Suzhou, China), 69–73. doi: 10.1145/3340555.3353744

Anjos, R. K. D., Sousa, M., Mendes, D., Medeiros, D., Billinghurst, M., Anslow, C., et al. (2019). "Adventures in hologram space: exploring the design space of eye-to-eye volumetric telepresence," in *25th ACM Symposium on Virtual Reality Software and Technology* (Parramatta, NSW), 1–5. doi: 10.1145/3359996.3364244

Argyle, M., and Cook, M. (1976). *Gaze and Mutual Gaze*. Cambridge: Cambridge University Press.

Argyle, M., and Dean, J. (1965). Eye-contact, distance and affiliation. *Sociometry* 289–304. doi: 10.2307/2786027

Bailenson, J. N., Swinth, K., Hoyt, C., Persky, S., Dimov, A., and Blascovich, J. (2005). The independent and interactive effects of embodied-agent appearance and behavior on self-report, cognitive, and behavioral markers of copresence in immersive virtual environments. *Presence* 14, 379–393. doi: 10.1162/105474605774785235

Baran, I., and Popović, J. (2007). Automatic rigging and animation of 3D characters. *ACM Trans. Graph*. 26, 72–es. doi: 10.1145/1276377.1276467

Bernard, I. (1997). *Film and Television Acting: From Stage to Screen*. Boca Raton, FL: CRC Press. doi: 10.4324/9780080506364

Braun, V., and Clarke, V. (2013). *Successful Qualitative Research: A Practical Guide for Beginners*. Newbury Park, CA: Sage Publications.

Bucher, J. (2017). *Storytelling for Virtual Reality: Methods and Principles for Crafting Immersive Narratives*. Taylor & Francis. doi: 10.4324/9781315210308

Budd, C., Huang, P., Klaudiny, M., and Hilton, A. (2013). Global non-rigid alignment of surface sequences. *Int. J. Comput. Vis*. 102, 256–270. doi: 10.1007/s11263-012-0553-4

Bye, K. (2017). *An Elemental Theory of Presence + Future of AI & Interactive Storytelling*. Available online at: https://voicesofvr.com/502-an-elemental-theory-of-presence-future-of-ai-interactive-storytelling (accessed May 28, 2020).

Cagniart, C., Boyer, E., and Ilic, S. (2010a). "Free-form mesh tracking: a patch-based approach," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (San Francisco, CA), 1339–1346. doi: 10.1109/CVPR.2010.5539814

Cagniart, C., Boyer, E., and Ilic, S. (2010b). "Probabilistic deformable surface tracking from multiple videos," in *European Conference on Computer Vision* (Crete: Springer), 326–339. doi: 10.1007/978-3-642-15561-1_24

Carranza, J., Theobalt, C., Magnor, M. A., and Seidel, H.-P. (2003). Free-viewpoint video of human actors. *ACM Trans. Graph*. 22, 569–577. doi: 10.1145/882262.882309

Casas, D., Tejera, M., Guillemaut, J.-Y., and Hilton, A. (2012). Interactive animation of 4D performance capture. *IEEE Trans. Visual. Comput. Graph*. 19, 762–773. doi: 10.1109/TVCG.2012.314

Chartrand, T. L., and Bargh, J. A. (1999). The chameleon effect: the perception-behavior link and social interaction. *J. Pers. Soc. Psychol*. 76:893. doi: 10.1037/0022-3514.76.6.893

Cho, S., Kim, S.-W., Lee, J., Ahn, J., and Han, J. (2020). "Effects of volumetric capture avatars on social presence in immersive virtual environments," in *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (Atlanta, GA), 26–34. doi: 10.1109/VR46266.2020.1581170537418

Collet, A., Chuang, M., Sweeney, P., Gillett, D., Evseev, D., Calabrese, D., et al. (2015). High-quality streamable free-viewpoint video. *ACM Trans. Graph*. 34:69. doi: 10.1145/2766945

Colyer, S. L., Evans, M., Cosker, D. P., and Salo, A. I. (2018). A review of the evolution of vision-based motion analysis and the integration of advanced computer vision methods towards developing a markerless system. *Sports Med. Open* 4:24. doi: 10.1186/s40798-018-0139-y

De Aguiar, E., Stoll, C., Theobalt, C., Ahmed, N., Seidel, H.-P., and Thrun, S. (2008). Performance capture from sparse multi-view video. *ACM Trans. Graph*. 27, 1–10. doi: 10.1145/1360612.1360697

De la Pe na, N., Weil, P., Llobera, J., Giannopoulos, E., Pomés, A., Spanlang, B., et al. (2010). Immersive journalism: immersive virtual reality for the first-person experience of news. *Presence* 19, 291–301. doi: 10.1162/PRES_a_00005

Debevec, P., Hawkins, T., Tchou, C., Duiker, H.-P., Sarokin, W., and Sagar, M. (2000). "Acquiring the reflectance field of a human face," in *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques* (New Orleans, LA), 145–156. doi: 10.1145/344779.344855

Depthkit (n.d.). *Depthkit*. Available online at: https://www.depthkit.tv/ (accessed May 14, 2020).

Dimension Studio (2019). *Wimbledon: Champion's Rally With Andy Murray*. Available online at: https://www.dimensionstudio.co/work/wimbledon-andy-murray (accessed May 15, 2020).

Dimension Studio (n.d.). *World Leading Volumetric, XR & Virtual Production Studio*. Available online at: https://www.dimensionstudio.co/ (accessed May 14, 2020).

Dodds, L. (2018). *Lighting a Legend: Art Pipelines in 'Hold The World' With Sir David Attenborough*. Available online at: https://www.gdcvault.com/play/1025611/Lighting-a-Legend-Art-Pipelines (accessed May 15, 2020).

Dolan, D., and Parets, M. (2016). *Redefining The Axiom Of Story: The VR And 360 Video Complex*. Available online at: https://techcrunch.com/2016/01/14/redefining-the-axiom-of-story-the-vr-and-360-video-complex/ (accessed August 30, 2020).

Eisert, P., and Hilsmann, A. (2020). "Hybrid human modeling: making volumetric video animatable," in *Real VR-Immersive Digital Reality*, eds M. Magnor and A. Sorkine-Hornung (Cham: Springer), 167–187. doi: 10.1007/978-3-030-41816-8_7

Garau, M. (2003). *The impact of avatar fidelity on social interaction in virtual environments* (Ph.D. thesis). University of London, London, United Kingdom.

Garau, M., Slater, M., Bee, S., and Sasse, M. A. (2001). "The impact of eye gaze on communication using humanoid avatars," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Seattle, WA), 309–316. doi: 10.1145/365024.365121

Gonzalez-Franco, M., Steed, A., Hoogendyk, S., and Ofek, E. (2020). Using facial animation to increase the enfacement illusion and avatar self-identification. *IEEE Trans. Visual. Comput. Graph*. 26, 2023–2029. doi: 10.1109/TVCG.2020.2973075

Gorisse, G., Christmann, O., Houzangbe, S., and Richir, S. (2019). From robot to virtual doppelganger: Impact of visual fidelity of avatars controlled in third-person perspective on embodiment and behavior in immersive virtual environments. *Front. Robot. AI* 6:8. doi: 10.3389/frobt.2019.00008

Gueguen, N., Jacob, C., and Martin, A. (2009). Mimicry in social interaction: its effect on human judgment and behavior. *Eur. J. Soc. Sci*. 8, 253–259.

Guo, K., Lincoln, P., Davidson, P., Busch, J., Yu, X., Whalen, M., et al. (2019). The relightables: volumetric performance capture of humans with realistic relighting. *ACM Trans. Graph*. 38, 1–19. doi: 10.1145/3355089.3356571

Hale, J., Ward, J. A., Buccheri, F., Oliver, D., and Hamilton, A. F. d. C. (2020). Are you on my wavelength? Interpersonal coordination in dyadic conversations. *J. Nonverb. Behav*. 44, 63–83. doi: 10.1007/s10919-019-00320-3

Hall, E. T. (1966). *The Hidden Dimension, Vol. 609*. Garden City, NY: Doubleday.

Hecker, C., Raabe, B., Enslow, R. W., DeWeese, J., Maynard, J., and van Prooijen, K. (2008). Real-time motion retargeting to highly varied user-created morphologies. *ACM Trans. Graph*. 27, 1–11. doi: 10.1145/1360612.1360626

Horry, Y., Anjyo, K.-I., and Arai, K. (1997). "Tour into the picture: using a spidery mesh interface to make animation from a single image," in *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques* (Los Angeles, CA), 225–232. doi: 10.1145/258734.258854

Huang, P., Budd, C., and Hilton, A. (2011). "Global temporal registration of multiple non-rigid surface sequences," in *CVPR 2011* (Colorado Springs, Co), 3473–3480. doi: 10.1109/CVPR.2011.5995438

Jensen, H. W., Marschner, S. R., Levoy, M., and Hanrahan, P. (2001). "A practical model for subsurface light transport," in *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques* (Vancouver, BC), 511–518. doi: 10.1145/383259.383319

Jump Studio (n.d.). *Jump Studio*. Available online at: https://www.jumpstudio.co.kr/ (accessed May 14, 2020).

Kanade, T., Rander, P., and Narayanan, P. (1997). Virtualized reality: constructing virtual worlds from real scenes. *IEEE Multimed*. 4, 34–47. doi: 10.1109/93.580394

Kätsyri, J., Förger, K., Mäkäräinen, M., and Takala, T. (2015). A review of empirical evidence on different uncanny valley hypotheses: support for perceptual mismatch as one road to the valley of eeriness. *Front. Psychol*. 6:390. doi: 10.3389/fpsyg.2015.00390

Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta Psychol*. 26, 22–63. doi: 10.1016/0001-6918(67)90005-4

Keskinen, T., Mäkelä, V., Kallionierni, P., Hakulinen, J., Karhu, J., Ronkainen, K., et al. (2019). "The effect of camera height, actor behavior, and viewer position on the user experience of 360 videos," in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (Osaka), 423–430. doi: 10.1109/VR.2019.8797843

Kovar, L., Gleicher, M., and Pighin, F. (2008). "Motion graphs," in *ACM SIGGRAPH 2008 Classes* (Los Angeles, CA), 1–10. doi: 10.1145/1401132.1401202

Levoy, M., and Hanrahan, P. (1996). "Light field rendering," in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques* (New Orleans, LA), 31–42. doi: 10.1145/237170.237199

Loper, M., Mahmood, N., Romero, J., Pons-Moll, G., and Black, M. J. (2015). SMPL: A skinned multi-person linear model. *ACM Trans. Graph*. 34:248. doi: 10.1145/2816795.2818013

MacQuarrie, A., and Steed, A. (2019). "Perception of volumetric characters' eye-gaze direction in head-mounted displays," in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (Osaka: IEEE), 645–654. doi: 10.1109/VR.2019.8797852

MacQuarrie, A., and Steed, A. (2020a). Free-viewpoint video. Available online at: https://vr.cs.ucl.ac.uk/portfolio-item/free-viewpoint-video/ (Accessed July 30, 2020).

MacQuarrie, A., and Steed, A. (2020b). "Improving free-viewpoint video content production using RGB-camera-based skeletal tracking," in *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (Atlanta, GA), 775–776. IEEE. doi: 10.1109/VRW50115.2020.00238

Mangiante, S., Klas, G., Navon, A., GuanHua, Z., Ran, J., and Silva, M. D. (2017). "VR is on the edge: how to deliver 360 videos in mobile networks," in *Proceedings of the Workshop on Virtual Reality and Augmented Reality Network* (Los Angeles, CA), 30–35. doi: 10.1145/3097895.3097901

Mathur, M. B., and Reichling, D. B. (2016). Navigating a social world with robot partners: a quantitative cartography of the uncanny valley. *Cognition* 146, 22–32. doi: 10.1016/j.cognition.2015.09.008

Metastage (n.d.). *Metastage*. Available online at: https://metastage.com/ (accessed May 14, 2020).

Mori, M. et al. (1970). The uncanny valley. *Energy* 7, 33–35.

Mustafa, A., Kim, H., and Hilton, A. (2016). "4D match trees for non-rigid surface alignment," in *European Conference on Computer Vision* (Amsterdam: Springer), 213–229. doi: 10.1007/978-3-319-46448-0_13

Narayanan, P., Rander, P. W., and Kanade, T. (1998). "Constructing virtual worlds using dense stereo," in *Sixth International Conference on Computer Vision (Bombay: IEEE Cat. No. 98CH36271)*, 3–10. doi: 10.1109/ICCV.1998.710694

Novick, D. G., Hansen, B., and Ward, K. (1996). "Coordinating turn-taking with gaze," in *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP'96, Vol. 3* (Philadelphia, PA), 1888–1891. doi: 10.1109/ICSLP.1996.608001

Oculus Story Studio (2015). *The Swayze Effect*. Available online at: https://www.oculus.com/story-studio/blog/the-swayze-effect/ (accessed May 28, 2020).

Prada, F., Kazhdan, M., Chuang, M., Collet, A., and Hoppe, H. (2016). Motion graphs for unstructured textured meshes. *ACM Trans. Graph*. 35, 1–14. doi: 10.1145/2897824.2925967

Pyett, P. M. (2003). Validation of qualitative research in the "real world". *Qual. Health Res*. 13, 1170–1179. doi: 10.1177/1049732303255686

Regateiro, J., Volino, M., and Hilton, A. (2018). "Hybrid skeleton driven surface registration for temporally consistent volumetric video," in *2018 International Conference on 3D Vision (3DV)* (Verona: IEEE), 514–522. doi: 10.1109/3DV.2018.00065

Roberts, D. J., Rae, J., Duckworth, T. W., Moore, C. M., and Aspin, R. (2013). Estimating the gaze of a virtuality human. *IEEE Trans. Visual. Comput. Graph*. 19, 681–690. doi: 10.1109/TVCG.2013.30

Rothe, S., Kegeles, B., and Hußmann, H. (2019). "Camera heights in cinematic virtual reality: How viewers perceive mismatches between camera and eye height," in *Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video* (Salford, UK), 25–34. doi: 10.1145/3317697.3323362

Salazar Kämpf, M., Liebermann, H., Kerschreiter, R., Krause, S., Nestler, S., and Schmukle, S. C. (2018). Disentangling the sources of mimicry: Social relations analyses of the link between mimicry and liking. *Psychol. Sci*. 29, 131–138. doi: 10.1177/0956797617727121

Schreer, O., Feldmann, I., Kauff, P., Eisert, P., Tatzelt, D., Hellge, C., et al. (2019). "Lessons learnt during one year of commercial volumetric video production," in *Proceedings of IBC Conference* (Amsterdam).

Sky UK Ltd (2018). *Sky VR: Hold the World*. Available online at: https://www.oculus.com/experiences/rift/2331434793563555/ (accessed May 15, 2020).

Slater, M., Gonzalez-Liencres, C., Haggard, P., Vinkers, C., Gregory-Clarke, R., Jelley, S., et al. (2020). The ethics of realism in virtual and augmented reality. *Front. Virtual Real*. 1:1. doi: 10.3389/frvir.2020.00001

Slater, M., Khanna, P., Mortensen, J., and Yu, I. (2009a). Visual realism enhances realistic response in an immersive virtual environment. *IEEE Comput. Graph. Appl*. 29, 76–84. doi: 10.1109/MCG.2009.55

Slater, M., Lotto, B., Arnold, M. M., and Sánchez-Vives, M. V. (2009b). How we experience immersive virtual environments: the concept of presence and its measurement. *Anuar. Psicol*. 40, 193–210.

Spanlang, B., Navarro, X., Normand, J.-M., Kishore, S., Pizarro, R., and Slater, M. (2013). "Real time whole body motion mapping for avatars and robots," in *Proceedings of the 19th ACM Symposium on Virtual Reality Software and Technology* (Singapore), 175–178. doi: 10.1145/2503713.2503747

Spanlang, B., Normand, J.-M., Borland, D., Kilteni, K., Giannopoulos, E., Pomés, A., et al. (2014). How to build an embodiment lab: achieving body representation illusions in virtual reality. *Front. Robot. AI* 1:9. doi: 10.3389/frobt.2014.00009

Starck, J., and Hilton, A. (2003). "Model-based multiple view reconstruction of people," in *IEEE International Conference on Computer Vision* (Nice), 915–922. doi: 10.1109/ICCV.2003.1238446

Starck, J., and Hilton, A. (2007). Surface capture for performance-based animation. *IEEE Comput. Graph. Appl*. 27, 21–31. doi: 10.1109/MCG.2007.68

Starck, J., Miller, G., and Hilton, A. (2005). "Video-based character animation," in *Proceedings of the 2005 ACM SIGGRAPH/Eurographics Symposium on Computer Animation* (Los Angeles, CA), 49–58. doi: 10.1145/1073368.1073375

Start, V. R. (2018). *Awake: Episode One*. Available online at: https://store.steampowered.com/app/845900/Awake_Episode_One/ (accessed May 15, 2020).

Steed, A., Pan, Y., Watson, Z., and Slater, M. (2018). "We wait"-the impact of character responsiveness and self embodiment on presence and interest in an immersive news experience. *Front. Robot. AI* 5:112. doi: 10.3389/frobt.2018.00112

Subramanyam, S., Li, J., Viola, I., and Cesar, P. (2020). "Comparing the quality of highly realistic digital humans in 3DoF and 6DoF: a volumetric video case study," in *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (Atlanta, GA), 127–136. doi: 10.1109/VR46266.2020.1581260728335

SuperData (2020). *2019 Year In Review Digital Games and Interactive Media*. Available online at: https://www.superdataresearch.com/2019-year-in-review (accessed May 16, 2020).

Tung, T., and Matsuyama, T. (2010). "Dynamic surface matching by geodesic mapping for 3D animation transfer," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (San Francisco, CA), 1402–1409. doi: 10.1109/CVPR.2010.5539806

Uva, M. (2010). *The Grip Book*. London, UK: Taylor & Francis.

Vinayagamoorthy, V., Brogni, A., Gillies, M., Slater, M., and Steed, A. (2004). "An investigation of presence response across variations in visual realism," in *The 7th Annual International Presence Workshop* (Valencia), 148–155.

Vlasic, D., Baran, I., Matusik, W., and Popović, J. (2008). Articulated mesh animation from multi-view silhouettes. *ACM Trans. Graph*. 27:97. doi: 10.1145/1360612.1360696

Volucap (n.d.). *Volucap*. Available online at: https://volucap.de/ (accessed May 14, 2020).

Xu, F., Liu, Y., Stoll, C., Tompkin, J., Bharaj, G., Dai, Q., et al. (2011). Video-based characters: creating new human performances from a multi-view video database. *ACM Trans. Graph*. 30:32. doi: 10.1145/2010324.1964927

Zhao, S. (2003). Toward a taxonomy of copresence. *Presence* 12, 445–455. doi: 10.1162/105474603322761261

Zitnick, C. L., Kang, S. B., Uyttendaele, M., Winder, S., and Szeliski, R. (2004). High-quality video view interpolation using a layered representation. *ACM Trans. Graph*. 23, 600–608. doi: 10.1145/1015706.1015766

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.